EVALUATING THE INFLUENCE OF SECOND SCREEN USAGE ON VIDEO QUALITY PERCEPTION

Jonas De Meulenaere[1a], Nicolas Staelens[2], Wendy Van den Broeck[1], Filip De Turck[2], Piet Demeester[2]

iMinds-SMIT, Vrije Universiteit Brussel, Belgium[1]
Pleinlaan 9, B -1050 Brussels, Belgium
Ghent University - iMinds, Department of Information Technology (INTEC), Ghent, Belgium[2]
Gaston Crommenlaan 8, B-9050 Ghent, Belgium
Jonas.de.meulenaere@vub.ac.be[a]

ABSTRACT

This study investigates to what extent subjects are influenced by second screen (2nd screen) usage in terms of perceiving and evaluating video quality impairments in an adaptive streaming scenario. A controlled between subjects test was set up to ascertain whether subjects that are using a related 2nd screen are less sensitive to video quality impairments compared to subjects not using a 2nd screen. Our results show a significant influence of 2nd screen usage in terms of impairment detection and evaluation. However, tests to measure whether this was attributable to a divided attention showed no effect.

1. INTRODUCTION

In this paper we report on research in which we investigate the Quality of Experience (QoE) in the context of simultaneous 2nd screen usage. As such, this research taps on two emerging trends: Over-The-Top (OTT) streaming in increasing video quality and combining multiple screens in a home context. Technical bottlenecks for OTT video streaming such as bandwidth and download limitations (in the case of Belgium) are being relaxed, allowing TV providers to offer customers such services as a side business (next to IPTV or cable).

Next, we also witness an increasing penetration of tablet computers and smartphones. In Flanders in September 2012 38,5% owned a smartphone and 27,7% owned a tablet (+14,6% compared to 2011). Both devices are regularly used for multitasking: 74.9% of tablet owners and 57,4% of mobile phone owners use their respective device while watching TV [2]. This established behaviour of combining the simultaneous usage of two devices provides broadcasters, content providers and content aggregators with the opportunity to address these devices as 2nd screen companions to the TV viewing experience.

2. BACKGROUND AND RELATED WORK

One of the solutions currently gaining ground for optimal OTT video delivery is HTTP Adaptive Streaming (HAS). HAS aims to take into account broadband network configurations as well as the video playback device and the transfered content. In that respect, it delivers video sequences in various bitrates (or other attributes) in order to match the available bandwidth and client characteristics, but can dynamically adapt to changes in this configuration [5]. Previous subjective tests [6,9] indicate that a constant bitrate is prefered to an oscillating one. Moreover, viewers prefer a lower drop in video quality rather than a constantly changing bitrate. Related to content features, viewers are more likely to detect quality variations in stills and slow pans than scenes with rapid motion [4, 6]. The underlying explanation is that the viewer's mind requires time to adjust to the scene before its video quality can be evaluated [6].

This explanation emphasises that evaluating video quality is a cognitive process, combining a low level perceptive and a high level reference path [3]. The quality perception path entails the physical perception of a stimulus, such as video images. The reference path functions as a constraint for this low level path as it reflects the previous quality experiences as well as the "*temporal and contextual nature of the quality formation process.*"[3]. Combined they form the quality evaluation of the perceived stimulus.

A low level human factor that influence QoE is for instance the level of attention while perceiving a stimulus. Arguably, the usage of a 2nd screen implies that viewers divide their visual attention over two screens, which may also imply a divided attention. A dual task can have a significant impact on how video quality is evaluated as is shown in previous work of [7]. In this subjective test interpreters were confronted with various audio-video (AV) desynchronizations while they had to translate the

sequence. Compared to subjects in a control group these interpreters had a significantly higher detection and annoyance treshold. They indicated that the translation task was cognitively too demanding to perceive the (minor) AV-desynchronizations, which emphasises the importance of the subject's primary focus (in this case the translating task).

In order to investigate the influence of 2nd screen usage in terms of video quality impairment detection and evaluation, we focus on the aspects of video quality acceptability, and impairment detection and evaluation. The main hypotheses we want to investigate are that (H1) subjects using a 2nd screen are less likely to detect video quality impairments and (H2) will rate an impaired video signal as more acceptable compared to subjects not using a 2nd screen. In addition, we will also investigate whether the extent to which a subject divides his/her attention between the TV and the tablet will (H3) influence the impairments detection and (H4) the tolerance towards the detected impairments.

3. SUBJECTIVE QUALITY ASSESSMENT

3.1. Experimental setup

To investigate the influence of video quality impairments in an adaptive streaming scenario a controlled subjective test was conducted, in which an experimental group was compared to a control group. For this test a 40" HDTV-set was used. This set was placed in a controlled test environment that resembled a living room. Subjects were seated at a distance of 235 cm of the TV set.

In line with previous research [8], we aim to use meaningful content for this test in order to increase the ecological validity of this test. A sequence of an existing cooking program was selected (Plat Préféré – broadcasted by the Flemish Public Service Broadcaster). The content is in 720p resolution and we selected a 320 seconds excerpt from it.

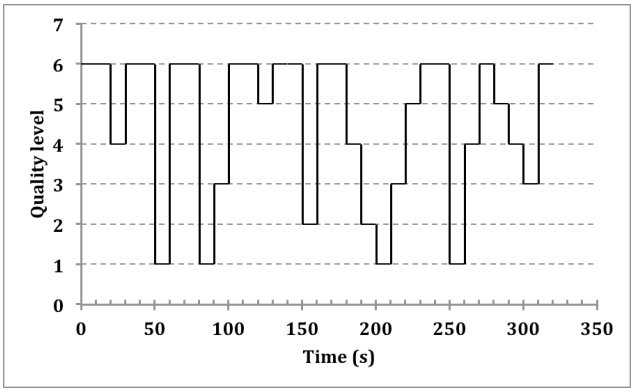

Figure 1 Adaptive streaming scenario

In order to simulate deployment over HTTP Live Streaming, the sequence was encoded following Apple's Technical Note TN2224 [1]. During play out, the quality fluctuates along 10s intervals, with 32 intervals in total (Figure 1). The sequence starts and ends with perfect quality; in between, the quality level fluctuates between best and worst quality in various order.

3.2 Protocol

Subjects were assigned randomly to either the experimental or the control condition. One subject at the time participated in this test. The test subjects were informed about the purpose of the experiment (i.e. video quality) and the nature of the impairments they might encounter in the test sequence. Test subjects in the experimental condition were handed a tablet computer (a fourth generation iPad) and instructed to perform an activity on the tablet during the test sequence (i.e. follow the recipe of the dish that is prepared in the sequence).

3.3 Used measures

After watching the test sequence and performing the task, all subjects were asked to answer the questions on the identified subjective measures. The perceived overall video quality was measured on a 5-point absolute category rating (ACR) scale, ranging from *bad* to *excellent*. Next, subjects had to indicate whether they had detected any impairments in the test sequence (yes or no) and, if yes, to what extent they considered these impairments to be annoying on a 5-point ACR scale, ranging from *very annoying* to *imperceptable* [10].

The subjects perceived attention was subjectively measured using two quasi continuous scales, one for perceived attention towards the tablet and one for the perceived attention towards TV. Two indicators were placed on the scales on both extremes: *little attention* and *much attention*. These scales were later divided in 25 intervals to include them in the analyses.

The extent to which subjects consider video quality important for TV-shows such as Plat Préféré was also measured using a 5-point scale (*not important at* all to very *important*). Besides these subjective measures also two objective measures were taken into account. The first being the allocation of the subjects' visual attention between 1st and 2nd screen, which was measured in duration. The second is the alternation of the subjects visual attention between the TV and the tablet. Both measures were logged using software that was operated by the researcher who observed the tests. The test ended with a short interview in which the research subjects discussed their experience in order to contextualise the findings.

3.4 Sample

Two groups of 16 subjects (total N=32) participated in the experiment, one under the experimental condition

(activity: follow the recipe) and one in the non-experimental condition (no activity). The sample consisted of 12 men and 20 women, which were more or less evenly distributed over both conditions (EC: 11F, 5M; CC: 9F, 7M). All of them can be considered non-experts.

4. RESULTS

4.1 Perceived overall video quality

The mean score for the perceived overall video quality was 3,75 for the experimental condition and 3,31 for the control condition. As the assumption of homogeneity was not met we performed a Mann-Whitney test. This test indicates that both conditions do not differ significantly in terms of overall video quality perception (Z=-1,237; p=0,216).

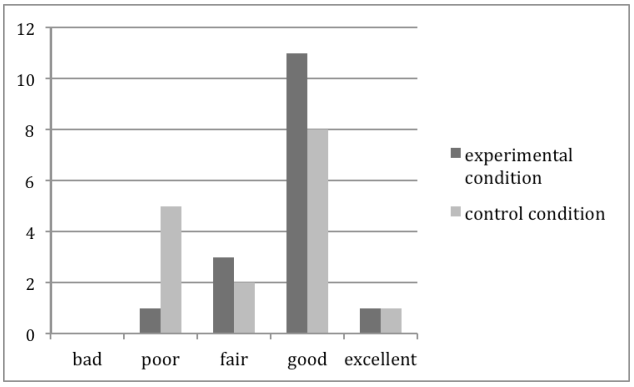

Figure 2 Perceived overall video quality (N=32)

In most instances, subjects indicated that the perceived errors were too short in duration and too limited to really affect their assessment of the entire sequence. These statements are supported by the low but significant positive correlation of overall quality evaluation and the evaluation of the perceived impairments (r(32)=0,321; p=0,073).

4.2 Detection rate

Six of the 16 subjects in the experimental condition indicated they did not notice any impairment. In the control group, only one subject did not notice any impairment. A Mann-Whitney test indicated that both conditions indeed differ significantly in terms of impairment detection (Z=-2,104; p=0,035) and subjects using a 2nd screen do detect less impairments compared to subjects only watching the TV-screen. This suggests that the tablet might distract viewers enough from potential quality impairments to detect them.

A limited cognitive processing capacity in this dual task context can be a first explanation (cf. [7]). Another explanation may reside in the fact that because of the switching visual attention subjects do not notice the transition of one visual quality level/layer to another. This would imply that these subjects perceive one continuous quality level [cf. 6, 9].

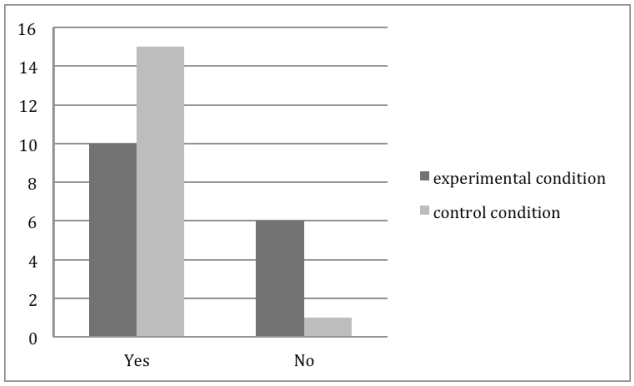

Figure 3 Impairment detection

4.3 Impairment annoyance

Although they were informed about what types of errors they might see in the test sequence, the types of impairments subjects detected ranged from degradations in resolutions over a shaky camera to content features itself (e.g. unintended appearance of film crew). On the one hand this indicates that the absolute evaluative scores should be considered carefully as non-video related quality issues might influence them. On the other hand this might imply that the simulated video quality impairments of the sequence are rather mild and require an informed subject to detect them.

The mean score for the evaluation of the perceived video quality impairments was 4,06 for the experimental condition and 3,25 for the control condition. We performed an ANOVA to test the significance of this difference and found that subjects using a 2nd screen do indeed tolerate the perceived video quality impairments more compared to subjects that are not using a 2nd screen (F(1, 30)=4,669; p= 0,039).

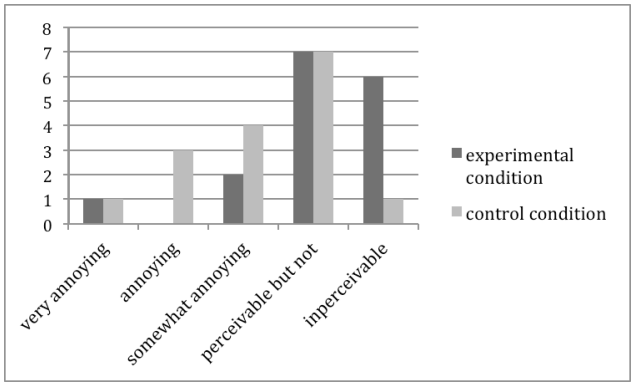

Figure 4 Impairment evaluation scores (N=32)

Notwithstanding the significant difference between both conditions, the evaluation scores are not mutually excluding. The subsequent interview indicates that there are differences in how impairments were perceived. Overall, the impairments were not considered that annoying, as the chart in Figure 4 indicates. However, subjects that consider the impairments very annoying, annoying or somewhat annoying particularly mention the fluctuating video quality as the most annoying factor to them, which is in line with previous research [6, 9]. It is not that other subjects did not perceive these fluctuations yet, these are considered to be not annoying and/or to be insignificant for this type of content to them.

4.4 Divided attention

In terms of the perceived division of attention subjects in the experimental condition indicated on average that they focused their attention more on the TV-set (M=20,44; SD=1,75) than on the tablet (M=7; SD =3,31). On average subjects spend 58,78 seconds watching the tablet (SD=35,81s) and 266,18 seconds watching the TV (SD = 35,88s).

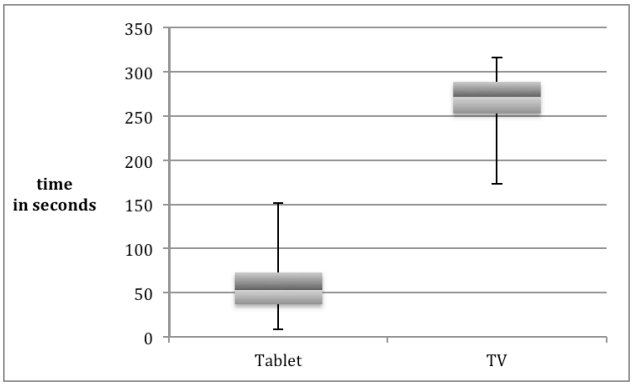

Figure 5 Tablet versus TV in terms of time spend watching (N=16)

This indicates that the tablet was really used as a secondary device. The tablet (with the recipe) was mainly used as a reference for the actions performed on the TV, and considered a side activity by all of the subjects. Moreover, in the subsequent interview most subjects indicated to be more interested in the content of the TV-screen than that of the tablet.

We also measured how many times subjects from the experimental condition switched their visual attention between tablet and TV and found that on average subjects switched 30,87 times, yet again again there was a large standard deviation (SD=22,14). This indicates that the tablet was used differently among the subjects. We noticed that some subjects hardly watched the tablet at all, whereas other subjects almost continuously switched their visual attention between both screens.

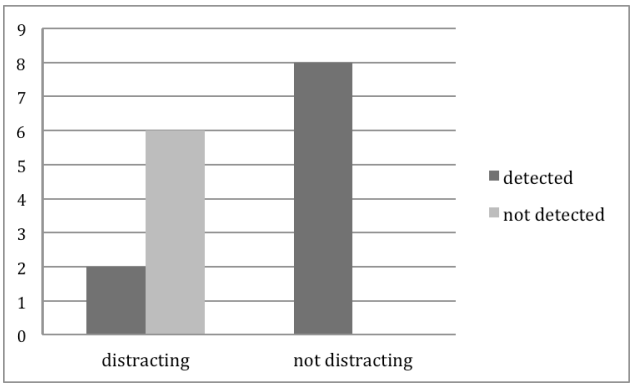

Figure 6 Detection rate related to state of being distracted by tablet (N=16)

The extent to which the tablet was considered distracting matches the impairment detection rate. Subjects that considered the tablet not distracting all indicated they detected impairments in the sequence. The majority of the subjects that did consider the tablet distracting, detected no quality impairments at all (see Figure 6).

We also wanted to test whether these attention measures (i.e. number of switches, gaze time per devices, and pereived attention) explained any of the impairment detection or evaluation. A Pearson correlation test indicated however no relation between any of the predictors. Therefore no further analyses were performed.

4.5 Importance of video quality

Subjects in both conditions were also asked to indicate to what extent they consider video quality important for TV-shows such as Plat Préféré or coocking programs in general. On average subjects across both conditions considered video quality in this context moderately important (M=3,78; SD=0,975). Subjects that consider the overall video quality as poor, indicate that video quality is important for this type (or any type) of AV content. In addition, optimal video quality is expected because they know it can be better and they are used to better quality. Moreover, a lower video quality might lower the appraisal of the content itself.

However only significant on a 0,1 level we found a low negative correlation (R(32)=-0,306; p=0,088) between the extent to which subjects considered video quality for this type of AV content important and how they evaluated the perceived impairments. Narrowed down to the experimental condition, no significant difference between the considered importance of video quality in terms of impairment detection was found.

5. DISCUSSION AND CONCLUSION

We performed a between subjects experiment to investigate whether the usage of a 2nd screen affects subjects' evaluation of the overal video quality of an

impaired sequence, as well as the detection rate and the evaluation of these impairments in a HAS scenario. Furthermore, we investigated whether these evaluations could be explained by a variation in attention.

We found significant differences in terms of impairment detection and impairment annoyance. However, this does not affect subjects' overall evaluation of the quality of the video sequence for neither of the two groups.

No connections between impairment detection and evaluation on the one hand and attention measures on the other hand were found. Possible explanations are inadequate parameters to measure the level of attention, insufficient statiscal power or attention is not the right parameter. The latter is rather unlikely, given the various accounts of the subjects that they felt distracted. Future research may focus on other ways to measure attention in this context.

Subjects were also asked to indicate to what extent they considered video quality important for this type of content. We found that the considered importance of video quality correlates significantly in a low negative way with the evaluation of the impairments. No difference was found, however, between subjects of the experimental condition that consider quality important and subjects that consider it unimportant in terms of evaluation of the impairments. Future research can determine whether next to the usage of a 2nd screen the cosidered importance of video quality for this type of AV content (or other variables) is also influencing the detection rate and evaluation score.

Our results show indeed that the simultaneous usage of 2nd screen (i.e. tablets) with content related to the 1st screen (=TV), lowers the detection rate of video quality impairments and suggests that 2nd screen users do tolerate more video quality impairments compared to non-2nd screen users. Given the significant quality oscilation and high amplitudes, our results diverge from previous tests with only focused viewers [cf. 6, 9]. Future research on this 2nd screen scenario is therefore advised to determine whether these findings hold or should be adapted.

ACKNOWLEDGEMENT

This research was done within the framework of the iMinds MISTRAL project. This project is co-funded by iMinds, a research institute founded by the Flemish Government. Organizations involved in the project include among others iMinds-SMIT, Vrije Universiteit Brussel and Ghent University - iMinds, Department of Information Technology (INTEC).

10. REFERENCES

[1] Apple, "Technical Note TN2224: Best Practices for Creating and Deploying HTTP Live Streaming Media for the iPhone and iPad," Available: http://developer.apple.com/library/ios/#tech-notes/tn2224/_index.html.

[2] iMinds-iLab.o, "Digimeter report 5, adoption and usage of media & ICT in Flanders wave 5 (Aug - Sept 2012)," iMinds, Gent, 2012.

[3] P. Le Callet, S. Möller, and P. Andrew, (eds.) "Qualinet White Paper on Definitions of Quality of Experience Output version of the Dagstuhl seminar." Dagstuhl, Jun, 2012.

[4] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality assessment of scalable video coding: a survey," in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, 2011, pp. 25–30.

[5] O. Oyman and S. Singh, "Quality of experience for HTTP adaptive streaming services," *Communications Magazine, IEEE*, vol. 50, no. 4, pp. 20–27, 2012.

[6] D. C. Robinson, Y. Jutras, and V. Craciun, "Subjective Video Quality Assessment of HTTP Adaptive Streaming Technologies," *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 5–23, Mar. 2012.

[7] N. Staelens, J. De Meulenaere, L. Bleumers, G. Wallendael, J. Cock, K. Geeraert, N. Vercammen, W. Van den Broeck, B. Vermeulen, R. Van de Walle, and P. Demeester, "Assessing the importance of audio/video synchronization for simultaneous translation of video sequences," *Multimedia Systems*, vol. 18, no. 6, pp. 445–457, May 2012.

[8] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "The Importance of Assessing Quality of Experience of IPTV and Video on Demand Services in Real-life Environments," 2010.

[9] M. Zink, J. Schmitt, and R. Steinmetz, "Layer-encoded video in scalable adaptive streaming," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 75–84, Feb. 2005.

[10] "Recommendation ITU-R BT.500-11 Methodology for the subjective assessment of the quality of television pictures."